

Evolving expectations and scholarly e-books

Based on a paper presented at the first UKSG one-day conference, *Usercentric: new strategies for scholarly communication*, London, November 2010

As they re-engineer workflows and rewrite the laws of publishing in the transition to online, scholarly publishers face great opportunities but also significant challenges, not least because users' expectations are formed by their experience of the internet giants with whom they interact daily. Publishers need to focus on both the discoverability and the delivery of their content to meet these expectations successfully; by leveraging human and machine networks they can enrich both their content and their users' experience of it through serendipitous connections. This paper focuses on Palgrave Macmillan's Palgrave Connect e-book platform¹, launched in 2009, as an example of such innovation.



ALISON JONES
Director of Digital Development
Palgrave Macmillan

I began my presentation to the UKSG conference in November 2010 with a slide showing images of the headquarters of internet giants including Google, Microsoft, Facebook, Yahoo!, Apple and Amazon. A recent Experian Hitwise study² indicated that the top ten sites accounted for a third (33%) of all US web traffic in 2010, which is astonishing given the sheer volume of websites (over 255 million at the end of 2010³).

My reason for choosing images of headquarters rather than simply company logos was to make a telling visual point: these technology companies have really nice offices. So very much nicer than those of the average scholarly publisher. The implication is clear: these are companies with deep, deep pockets and patient investors, whereas scholarly publishers have to make their investments out of tight operating margins and, in most cases, have to show a rapid return on investment to remain viable.

Sadly for scholarly publishers, though, our users do not make this distinction. When they visit our sites, and whether we like it or not, they cannot help but bring with them their aggregated experience of these internet giants. They expect intuitive interfaces, access from whatever device they happen to have to hand, with display optimized for that device, and connections that will take them from what they already know, or what they

know they want to learn, to what they don't know they don't know.

Figure 1 shows a version of the Johari window: useful on almost any occasion. Here, I propose it as a framework for the range of ways that users interact with information online: any site that delivers information needs strategies for each of these data-seeking modes. So, for example, if a user is starting in the least challenging area, reading about something they already know a good deal about for fun or pleasure, then a well-designed site can move them to explore less familiar territory. Similarly, if they come with a well-defined query, in exploration mode, we can

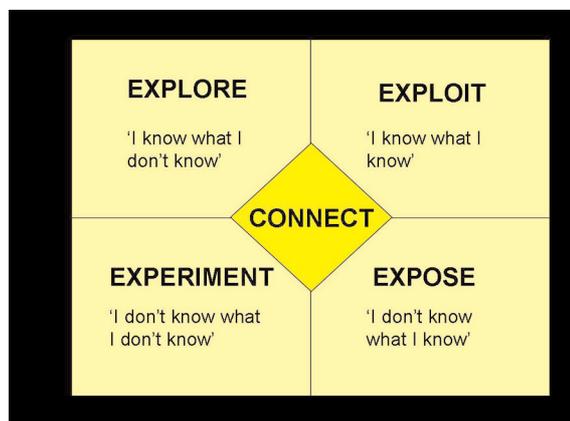


Figure 1. The Johari window

take them on to experiment with ideas that are entirely new to them, if we can get the connections right.

For the user these connections are purely conceptual and instinctive. It is our job as publishers to put the plumbing in place to facilitate them, ensuring the pipes are invisible as far as possible to the user. The two key components for these connections are: first, discoverability – how we enable users to find our content – and second, delivery – what we allow them to do with it. Digital monographs are increasingly taking on the characteristics of their short-form cousins, journal articles, and many of these strategies reflect this convergence.

Discoverability

Search

At the most basic level, our full text needs to be optimized for search engine discovery simply because 90% of students (and probably a similar proportion of scholars) start their search with Google. The Google discovery algorithm is notoriously opaque, but at the very least we can put content outside firewalls wherever possible, make that content dynamic and extensively linked, and provide as much meaningful metadata as possible.

For library resources such as Palgrave Connect, discoverability also means participating in federated search tools such as Primo Central, MetaLib, Summon and so forth. If publishers fail to include their content in the search tools that libraries use then to a large extent it does not matter how excellent the quality of that content, it is not going to be discovered, it is not going to be used and subscriptions will fall away because decisions to renew are based on usage. This is just a painful law of the jungle. We at Palgrave might believe that the Palgrave Connect interface offers a superior search experience but unfortunately there is a vanishingly small proportion of students and scholars prepared to make the trek direct to our site to discover that for themselves. Similarly, we are also included in link resolvers such as SFX so that users in subscribing institutions can go straight from reference to content. And we continually evaluate all the providers and the new tools as they emerge, to ensure that our content is as discoverable as we can make it for libraries all around the world.

Library catalogues

The library OPAC remains a key discovery tool for scholars and students, so we provide free MARC records to populate the catalogue and link to our content. This means working with OCLC for WorldCat subscribers, but for others we provide records from Nielsen. We seemed to spend a disproportionate amount of time on this in the early days of Palgrave Connect: we discovered there are subtle flavours of standards, such that a perfectly valid MARC record does not meet a particular institution's preferences. It is all very well encouraging publishers to adopt standards, but too often those standards are not as standard as we would like.

Network effects

Network discoverability effects operate on both human and machine levels and in both cases the fundamental law of scale holds true: the bigger, the better. The more data and the more users we have, the more powerful the effect.

At the human level, we are keen to harness the horizontal trust networks that thrive so well on the web. Obviously, there are ubiquitous referral services like Digg and Delicious that anybody can use on any site but we've also implemented the more specialist scholarly bookmarking tool, Connotea, because we know that recommendation from somebody who is a trusted person in your network is worth thousands of pounds of marketing spend on our part.

At the machine level, there are several important elements to our use of aggregated usage data. At the most basic level, we can analyse patterns to discover, perhaps, where people tend to leave the site, to identify areas where navigation can be improved. More interestingly, we can make transparent the connections our users have created by their paths through our site, and bring to our users' attention particularly popular titles, or (à la Amazon) indicate other titles a reader might be interested in, based on other users' viewing history.

Delivery

Access

At its most basic, providing access to subscribers means working with existing access control

systems in use by institutions: Shibboleth, Athens, IP recognition, referral from a secure site, and so on. It also means compliance with accessibility standards and we worked closely with JISC TechDis to ensure that Palgrave Connect is compliant with all the accessibility technologies that students with learning difficulties or disabilities are using to support their learning.

Interface

Do you remember the first time you saw the Google home page? For me, it was on a visit to a librarian in October 1999, when most websites looked like very cluttered magazine pages. Google's audacious, spacious design blew me away. Palgrave Connect does not have quite that minimalist purity, but we have put a lot of energy into ensuring we keep the clutter under control, that we exploit heuristics that people are familiar with from other sites, we minimize the burden of a new interface as far as possible. As Einstein said, things should be made as simple as possible, but no simpler.

Digital rights management

Another aspect of delivery is the vexed question of digital rights management (DRM), known to its proponents as essential protection of intellectual property, and to its detractors as crippleware that penalizes legitimate consumers. Palgrave Connect employs no technical DRM, only a watermark generated on the fly which identifies the institution and session of the download and enables us to track any abuse. Librarians are sometimes frustrated by publishers' inexplicable desire to keep control of their content rather than just popping it into a big aggregated pot with a single interface, which is one of our reasons for this decision. Where we own the content, where we have visibility of how it is being used and a direct relationship with our customers, we can balance our obligations to protect our authors' intellectual property with our responsibility to our users, and dispense with technical shackles. Interestingly, when we first decided to dispense with technical DRM, it felt like quite a risk, but it was key to the user experience we wanted to create, and so far our faith in our users has been rewarded. There has been no substantial abuse or piracy. However, it is fair to say that different types of content merit different approaches to DRM: whereas scholars and librarians tend to respect intellectual property, and limited

piracy is unlikely to have a significant effect on what is primarily a library purchasing base, textbooks, for example, would be a very different case. It just takes one student (or indeed instructor) in each institution sharing their file with a class of 200 students and suddenly the economics are unsustainable. To paraphrase Einstein, DRM should be as light as possible but no lighter.

Ubiquity

We want our content to be consumed wherever the user finds himself, with whatever device he has to hand, but we also want to record the usage (since, as we have already established, librarians make their renewal decisions based on usage statistics). Since our site has to function well on a range of mobile interfaces, in addition to PDFs we have begun making available ePub files, which allow text to reflow on mobile devices for a better reading experience.

Immediacy

Not only does the page need to load quickly (although of course this is also true), but librarians are no longer willing to tolerate a delay between print and electronic publication. To be fair to publishers, this delay was not due to a simple desire to irritate librarians on our part; we have had to re-work our production processes quite extensively. Palgrave have now completed that process – and most publishers are catching up if they are not there already – with titles now uploaded monthly onto Palgrave Connect as they are published.

Connectedness

At the risk of disappearing into a self-referential fog, 'connectedness' – to users, to related resources, to the whole big, wide world of scholarship – is not only the basis of the product's name, but its core philosophy and the focus of most of our ongoing development. For example, we have just introduced 'semantic fingerprinting' of each title, enabled by Palgrave Connect's powerful MarkLogic back-end. Effectively we data mine the full text of over 7,000 titles to enable us to identify those with similar semantic patterns. Such connections would almost certainly never be identified manually by editorial or marketing staff, which is both a strength and a weakness. We try to balance this by providing results on two levels: the first set of matches is constrained by subject area, so the matches are

pretty much guaranteed to be highly relevant, but the second set ranges across the whole platform and typically produces unexpected cross-disciplinary connections. Such serendipity in multiple dimensions is quite exciting, I hope you'll agree.

The distinctions between traditionally separate formats of scholarly publishing are becoming increasingly irrelevant for most users. If you are looking for scholarly research in a particular area, then you probably don't really care whether it calls itself a journal article or a monograph, so we are currently working on providing search results that connect these traditional silos. (More broadly, and this is more properly the subject of a whole paper of its own, we are engaged, along with many other academic publishers, in re-engineering book content to take advantage of the mature online journal ecosystem for monograph content: abstracts, DOIs, inbound and outbound OpenURLs, citation indexes, etc. It is an exciting time in monograph publishing.)

And there is also our own connectedness to our customers. For publishers, it is incredibly interesting to have visibility of who is buying our content and how they are using it, and we use that feedback to develop the platform and the content, to feed back to the authors, to inform our commissioning strategy. It is a virtuous circle for us, and I hope for our users as well.

And finally...

Finally, alongside discoverability and delivery, there is a third D: Dull but Essential. This is the plumbing, the stuff the user does not have to think about but we really must. To make the supply chain and the discoverability of content work, we need systems and standards that allow us to pump out our metadata and content as painlessly and as frequently as possible. That means ONIX (at least 2.1) for booksellers, MARC for libraries. And for content, ePub is now well established as the standard format for mobile devices, and it is only going to get more powerful and ubiquitous once version 3.0 is released in 2011.

There is no point pumping out metadata efficiently if that data is garbage. So, right through the workflow – editorial, marketing, rights,

production – everyone involved in the publishing processes understands the importance of inputting data immediately, keeping it up to date, making sure it is good quality. Traditionally, publishers focused on their metadata at a single point of time, just ahead of the catalogue being produced: now it is daily, the feeds are continuous.

And of course there is no point in the publisher having high-quality data that is pumped out efficiently if the content cannot be sold. This is not so much an issue for new titles – most publishers now have their legal house in order and new contracts provide for a wide range of digital rights – but there is a great deal of value in their backlists which may be inaccessible if rights are not in place. Palgrave has made a major investment in going back to authors over the last ten years to seek the rights that we need to make their content available electronically. The result is that a vast amount of invaluable content from our backlist is now available on Connect. And, whereas print monograph sales traditionally peak on publication and tail off very fast (following their promotion and stock lifecycle), for online users, and particularly in the humanities and social sciences, publication date is often irrelevant. It is a phenomenon identified back in 2007 by Michael Jensen as 'the Deep Niche'⁴: the content itself, not the publishing lifecycle, drives use. The unlocking of the deep backlist, particularly for humanities, has been one of the unexpected benefits of the whole e-book revolution in monographs.

References

1. www.palgraveconnect.com (accessed 26 January 2011).
2. <http://www.hitwise.com/us/press-center/press-releases/facebook-was-the-top-search-term-in-2010-for-sec/> (accessed 6 January 2011).
3. <http://news.netcraft.com/archives/category/web-server-survey/> (accessed 6 January 2011).
4. Jensen, M, The Deep Niche, *Journal of Electronic Publishing*, 2007, 10(2) .

Article © Alison Jones

■ Alison Jones

Director of Digital Development

Palgrave Macmillan

Houndmills

Basingstoke

Hampshire RG21 6XS, UK

Tel (direct): +44 (0)1256 302622

E-mail: Alison.Jones@palgrave.com

www.palgrave.com

To view the original copy of this article, published in *Serials*, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=24&issue=1&spage=12>

The DOI for this article is 10.1629/2412. Click here to access via DOI:

<http://dx.doi.org/10.1629/2412>

For a link to the full table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=24&issue=1>